Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs)

Cheng Lu[†], Dong-Hoon Jeong[†], Karthik Kulkarni[†], Manoj Pillay[†], Kan Nobuta[†], Rana German[†], Shawn R. Thatcher[†], Christopher Maher[‡§], Lifang Zhang[‡], Doreen Ware[‡¶], Bin Liu[‖], Xiaofeng Cao[‖], Blake C. Meyers[†], and Pamela J. Green[†,††]

[†]Department of Plant and Soil Sciences and Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711; [‡]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; [§]Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794; [¶]North Atlantic Area Plant, Soil, and Nutrition Laboratory Research Unit, U.S. Department of Agriculture/Agricultural Research Service, Ithaca, NY 15853; and [‖]National Key Laboratory of Plant Genomics and Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100080, China

Edited by Xiaoyu Zhang, University of Georgia, Athens, GA, and accepted by the Editorial Board January 15, 2008 (received for review September 14, 2007)

Small RNAs (21–24 nt) are involved in gene regulation through translation inhibition, mRNA cleavage, or directing chromatin modifications. In rice, currently ≈240 microRNAs (miRNAs) have been annotated. We sequenced more than four million small RNAs from rice and identified another 24 miRNA genes. Among these, we found a unique class of miRNAs that derive from natural *cis*-antisense transcript pairs. This configuration generates miRNAs that can perfectly match their targets. We provide evidence that the miRNAs function by inducing mRNA cleavage in the middle of their complementary site. Their production requires Dicer-like 1 (DCL1) activity, which is essential for canonical miRNA biogenesis. All of the natural antisense miRNAs (nat-miRNAs) identified in this study have large introns in their precursors that appear critical for nat-miRNA evolution and for the formation of functional miRNA loci. These findings suggest that other natural *cis*-antisense loci with similar exon–intron arrangements could be another source of miRNA genes.

high-throughput sequencing | siRNA | small RNA | massively parallel signature sequencing (MPSS)

Most eukaryotes possess small RNA-based gene silencing systems that can down-regulate genes at transcriptional and posttranscriptional levels (1, 2). At least five classes of these small regulatory RNAs (21–24 nt) have been characterized, including microRNAs (miRNAs), heterochromatic siRNAs, transacting siRNAs (ta-siRNAs), natural antisense siRNAs (nat-siRNAs), and, in metazoans, the Piwi-interacting RNAs (3–7). miRNAs are processed from self-complementary transcripts by the activity of Dicer ribonucleases. siRNAs originate from longer, double-stranded (ds)RNA molecules and usually represent both strands of the RNA, although they are similar in biochemical structure to miRNAs and have some functional similarities. In plants, siRNAs typically derive from transposons, repetitive sequences, and transgenes. These siRNAs could be involved in DNA methylation and histone modifications that silence target transcription (8). Although nat-siRNAs also have been identified (9, 10), natural antisense miRNAs (nat-miRNAs) have not been reported in any system.

Plant miRNAs have near-perfect pairing to their targets and therefore generally cause mRNA cleavage. Numerous studies have demonstrated the critical role of miRNAs in controlling developmental processes and organ identity. As of April 2007, the miRNA Sanger database contained 916 plant miRNAs. The list is rapidly growing as a result of new deep-sequencing technologies for small RNA discovery. In *Arabidopsis*, small RNAs from various mutants, tissues, and developmental stages have been analyzed by high-throughput pyrosequencing (11–15). These efforts identified at least 184 miRNAs (≈70 families) in *Arabidopsis*. Cloning of miRNAs from lower plants such as moss indicates that some miRNAs are conserved over a long evolutionary distance. In fact, most miRNAs identified in the early studies (21/28) are conserved in more than one plant species, although some miRNAs (miR158, miR161, miR163, and miR173) are not well conserved (2). Extensive cloning and sequencing data have provided important insights into the conservation and evolution of plant small RNAs, suggesting that the discovery of conserved miRNAs has reached a plateau. Still, many nonconserved or narrowly conserved miRNAs remain to be identified (11, 15), and among these may be new types of miRNA genes.

Rice (*Oryza sativa*) is a primary source of food for more than half of the world's population and is considered the most important crop. Molecular and genetic resources for rice have grown significantly in recent years. Initially, 138 miRNAs representing 20 families were annotated as rice miRNA genes, primarily based on sequence conservation with *Arabidopsis* (16). More recently, new miRNAs have been revealed in rice by direct cloning and traditional sequencing (17–19). Because of the low depth of available rice small RNA data, many of these miRNAs have only been sequenced once, have been not vigorously tested, and have probably been misannotated. A more complete and unbiased view of the small RNA transcriptome can be achieved by the recent advances in next-generation sequencing technology.

Here, we applied high-throughput sequencing to identify miRNAs that have thus far proven difficult to find by using traditional cloning or *in silico* predictions. We sequenced more than four million small RNAs from six rice samples. Our data provided insight into the veracity of many prior miRNA annotations and identified 24 previously uncharacterized miRNAs, many specific to rice. Moreover, we uncovered a group of nat-miRNAs, which originate from the natural antisense strand of target genes. Processing of introns from the overlapping primary miRNA (pri-miRNA) transcripts generates a hairpin structure that can be further processed by Dicer-like 1 (DCL1). The mature nat-miRNAs typically cause mRNA cleavage in the middle of the complementary site. These nat-miRNAs and the possession of introns in their precursors are highly conserved among monocots. Our findings suggest an additional pathway for miRNA evolution, biogenesis, and function.

Author contributions: C.L., D.-H.J., B.C.M., and P.J.G. designed research; C.L., D.-H.J., R.G., and S.R.T. performed research; C.M., L.Z., D.W., B.L., and X.C. contributed new reagents/analytic tools; C.L., D.-H.J., K.K., M.P., K.N., B.C.M., and P.J.G. analyzed data; and C.L., B.C.M., and P.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. X.Z. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE10523).

[††]To whom correspondence should be addressed. E-mail: green@dbi.udel.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0708743105/DC1.

Results and Discussion

Known Rice miRNAs Are Well Represented in Massively Parallel Signature Sequencing (MPSS) Libraries. To investigate the miRNA component of the small RNA transcriptome in rice, six small RNA libraries were constructed from wild-type rice inflorescence, seedling, and stem tissues (20) and from seedlings treated with abscisic acid (ABA) or with the rice blast pathogen *Magnaporthe grisea*. MPSS sequencing generated 4,473,864 17-nt signatures from these six libraries [supporting information (SI) Table 2]. The abundance of each sequence in the libraries was normalized based on the relative cloning frequency in each library. All signatures were matched against the rice genome, full-length miRNA stem-loops as annotated in the Sanger miRNA database, and a set of annotated rRNA, tRNA, and other noncoding RNAs. Sequences overlapping precursors of previously identified miRNAs constitute ≈23% of the library. As in *Arabidopsis*, miRNAs also were the most abundant sequences identified within each library (SI Table 3) (11, 14, 15). We also found a few loci that generate phased small RNAs as discussed in *SI Results and Discussion* and SI Table 4. Many remaining small RNAs corresponded to repeat regions such as transposons, retroelements or simple repeats. Of the sequences, 175,428 (57%) were detected only once, suggesting that sequencing was not saturated.

In rice, 242 miRNAs representing ≈60 families have been cloned or predicted. All of the miRNA families conserved in *Arabidopsis* were detected in our libraries, with abundances as high as 80,594 transcripts per quarter million (TPQ) (miR168 in the MPSS SNM library). Generally, miRNA loci generate fewer small RNA species than siRNA loci. However, in our dataset, sequences with low abundance often were detected that matched to sequences other than the miRNA portion of the pre-miRNA transcripts. These sequences had start positions between −2 and +2 nt away from annotated 5′ ends, possibly because of the inaccuracy of DCL1 processing (15). Generally, signatures perfectly matching annotated miRNAs were the most highly abundant sequences identified from that locus. However, for three miRNAs, the position of the most abundant signature was shifted from the annotated miRNA by 1 or 2 nt. For example, the annotated miRNA156 family was detected in our dataset, but more weakly than a (+1) variant. Intriguingly, miR168 was the most abundant small RNA in all of the rice libraries (SI Table 3) in contrast to *Arabidopsis* (13). In *Arabidopsis*, the levels of miR168 and AGO1 may be fine-tuned by an autoregulatory loop (21). Although it is unclear whether an autoregulatory loop operates similarly in rice, the extraordinarily high level of miR168 in rice and other monocots (data not shown) suggests strict regulation of the AGO1 mRNA.

Many known miRNA families have multiple paralogous loci. Within each family, closely related members sometimes produce identical miRNAs, confounding the determination of which loci are expressed. Although studies have indicated that almost all *Arabidopsis* miRNA genes are expressed (22), little is known about rice. Our small RNA data provided an opportunity to address this deficiency. Although the mature miRNA sequences frequently mapped to multiple locations in the genome, the full miRNA precursor sequences diverge among family members. We confirmed the expression of 77 loci by uniquely mapping small RNAs to specific miRNA precursors. For example, the miR160 family has six members in rice. We found unique signatures specifically generated from miR160b–d and miR160f (SI Fig. 5). Evidence for the expression of two other members (miR160a and miR160e) is not definitive because only duplicate signatures were detected from these loci (SI Fig. 5).

Given the complexity of the rice small RNA population and preponderance of siRNAs, it is challenging to confidently identify nonconserved miRNAs. The conserved miRNA families were further validated by our MPSS data, because for many of them both miRNA and miRNA* species were detected. However, other rice miRNAs were not supported. Other than the 138 genes representing the ≈20 conserved and validated miRNA families (those in the range of miR156 to miR399 and miR528 to miR535), additional candidates have been proposed, named, and annotated but were poorly supported by our data (designations between miR408 and miR821). None of those candidates predicted solely by computational methods were found in our data (miR413 to miR426) (23). There are 95 other nonconserved rice miRNAs in miRBase, all cloned from size-fractionated cDNA libraries, predicted to form hairpin structures with flanking sequences, and many gave positive signals on RNA gel blots (18, 19). However, on careful examination, only 13 of the 95 had strong miRNA-like characteristics (miR408, miR435, miR437, miR440, miR444, miR528, miR529, miR530, miR531, miR535, miR810, and miR820a–c). The other 82 sequences are either borderline candidates or originate from siRNA-like clusters. For example, miR442 is one of many small RNAs generated from both strands and randomly distributed along a long inverted repeat (SI Fig. 6). Part of the repeat, possibly a miniature inverted-repeat transposable element (MITE), is highly duplicated in rice. Our observation argues that high-throughput, deep-sequencing data are perhaps essential for effectively and confidently annotating nonconserved miRNA genes.

Identification of Previously Uncharacterized miRNAs in Rice. Several computational filters were used to identify previously uncharacterized miRNAs (Fig. 1*A*). The filters were designed based on consensus properties of a reference set of known rice miRNAs (miR156 to miR408). Among the 21 miRNA families, 19 were detected in at least three libraries. miRNAs were predominantly the most abundant signatures in each library, with all known miRNA families but miR395 and miR397 present at >150 TPQ (SI Table 3). Next, we applied the following criteria to our six libraries (Fig. 1*A*). We removed signatures matching to rRNAs, tRNAs, and other structural noncoding RNAs. To minimize noise, we also removed signatures found in only one or two libraries (Fig. 1*A*). Small RNAs that were expressed at very low levels were excluded (sum of abundance <100 TPQ) as were those matching to >20 genomic locations and those originating from loci that generated siRNA-like clusters (high-density cluster small RNAs originate from both strands). Using this combined set of filters, we identified a group of just 39 signatures (of 381,330 distinct sequences) from 31 clusters (Fig. 1*A*). Candidates from each of the 31 loci were subjected to a manual folding analysis. We found that 24 loci (representing 16 families) potentially form miRNA precursor-like stem-loop structures, in which no more than six positions within the putative miRNA–miRNA* duplex region were unpaired (Table 1, SI Fig. 7). Detection of miRNA* sequences could indicate that the predicted fold-back is a DCL1 substrate. Generally, miRNA* signatures have much lower abundances than corresponding miRNAs. For most of our miRNA candidates, miRNA* species (or ±1-nt variants) were detected in our libraries (Table 1).

The 16 candidate miRNAs with miRNA*s were subjected to RNA gel confirmation (see Figs. 1*B* and 3*B*). We also used RNA gel blots to confirm small RNAs from two loci that were predicted to form stem-loop structures but lacked miRNA* species. Bands with a size of ≈21 nt were observed for 16 of them (Fig. 1*B*, Table 1). This high confirmation rate indicates that our computational filters are very strict and that there are many more uncharacterized miRNAs in rice.

None of these identified miRNAs have obvious orthologs in the *Arabidopsis* or *Populus trichocarpa* (poplar) genomes, based on sequence/hairpin similarity and target conservation. In fact, for most of these miRNAs, no orthologs were identifiable in any other species among the available plant genomic and expressed sequence tag (EST) sequences. Only one group of these miRNAs was conserved in monocot plants such as maize, sorghum, wheat, switchgrass, and Brachypodium (Table 1 and Fig. 2). All members of this set of miRNAs are predicted to target mRNAs for MADS

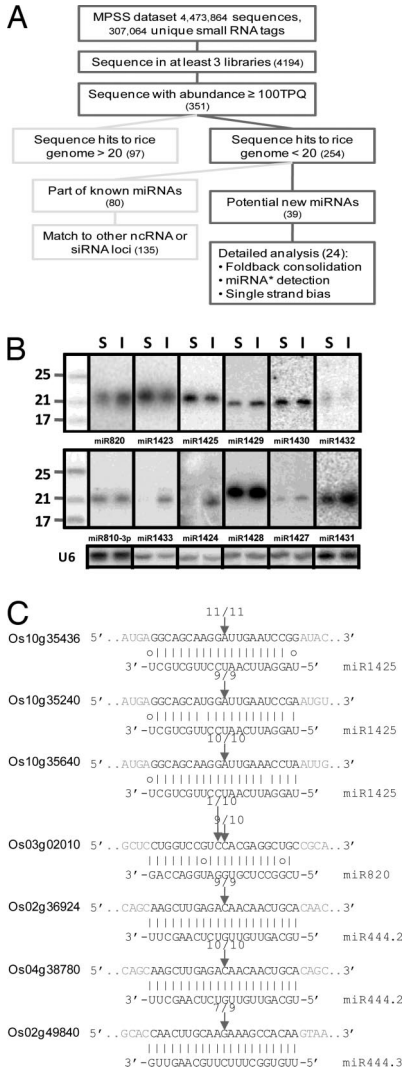

Fig. 1. Identification and analysis of rice miRNAs and targets. (*A*) Flowchart for the prediction of potential miRNAs from rice MPSS data. (*B*) RNA gel blots of LMW RNA isolated from inflorescence tissues (I) and 2-week-old seedlings (S) probed with labeled oligonucleotides. U6 loading controls are shown in the bottom row. (*C*) Validation of predicted targets. The arrow indicates a site verified by 5′ RACE, with the number of cloned RACE products sequenced shown above.

box proteins (Table 1 and Fig. 1*C*). Consistent with the *in silico* analysis, this group of miRNAs was detected by RNA gel blots in all of the tested monocot plants, but not in dicots, whereas the other previously uncharacterized miRNAs were observed only in rice (Fig. 2). In another independent analysis, we did not find additional, good, conserved miRNA candidates by comparing rice MPSS data with extensive maize small RNA data (our unpublished data).

Several genome-wide small RNA profiling projects revealed that a large number of nonconserved miRNA genes exist in *Arabidopsis* (11, 15). Most of these nonconserved miRNAs are represented by a single copy gene, consistent with a recent evolutionary origin (11, 24). Our results suggest that most of the identified rice miRNAs emerged after the divergence of rice and maize (≈50 million years ago). Given that nonconserved miRNAs usually are expressed at low levels, sometimes at very specific developmental stages or cell types, and that our analysis is focused on candidates that can be confirmed by RNA blots, there are probably many more to be discovered. The existence of a large number of species-specific miRNAs supports the notion that a diverse set of miRNA loci is evolving rapidly and independently within each species. Moreover, our sequencing effort uncovered a previously uncharacterized class of miRNAs (nat-miRNAs discussed below) that are highly conserved among monocots (Fig. 2).

Predicted Targets of Identified Rice miRNAs. To identify the potential targets of the identified miRNAs, we used a modified prediction algorithm based on additive, position-dependent mismatch penalties and compared against both the whole cDNA and genome sequence databases (16). Most conserved miRNA targets had a score <3.0 with the default setting (a score of zero is a perfect match), consistent with previously described miRNAs (16). Although some transcripts with a score of 3.5 can be validated by using a standard 5′ RACE analysis to detect cleavage location (11), a highly stringent score (≤3.0) was applied in our target prediction to maintain specificity and low noise. Targets were predicted with a penalty score of 3.0 or better for all of the identified miRNAs (Table 1), using the 21-nt sequences derived from the 17-nt MPSS tag plus four adjacent nucleotides from the matching genomic location. Some of highest-scoring potential targets were experimentally evaluated by using a modified 5′ RACE approach (Fig. 1*C*). From prior studies in *Arabidopsis*, predicted targets for nonconserved miRNAs have a lower confirmation rate (11). Possible reasons include low and/or different target expression, low abundance and/or activity of nonconserved miRNAs, or incorrect predictions (11).

The majority of conserved miRNA targets encode transcriptional factors that are key regulators in controlling developmental processes. However, nonconserved miRNA targets have a much broader spectrum of potential functions (11, 13, 15). Among the identified rice potential and validated miRNA targets, some resemble targets of previously identified miRNAs (Table 1). We confirmed that a DNA cytosine methyltransferase mRNA (Zmet3-like) was the target for osa-miR820, whereas a PPR domain-containing protein mRNA (Os02g03930) was predicted for osa-miR810-3p. Two putative transcriptional factor mRNAs (Os03g43390 and Os10g02970) were the predicted targets for osa-miR1428 and osa-miR1431, respectively. Several unique target families are worth noting. First, osa-miR1425 was determined to target the mRNAs of the fertility restorer 1 (RF1) gene family that contains pentatricopeptide repeats. RF1 encodes a mitochondrially targeted protein that has the ability to reduce the expression of the cytoplasmic male sterility-associated mitochondrial gene (25, 26). Second, osa-miR1432 was predicted to target at least three EF-hand family protein mRNAs. The EF-hand domain is a conserved helix–loop–helix structure that can bind a single Ca^{2+} molecule. Ca^{2+}-mediated signaling transduction pathways are widely used in plants (27). Our analysis suggests that miRNAs target mRNAs for proteins that are likely to play an important role in Ca^{2+} action. Overall, these targets broadened the biological processes regulated by miRNAs.

Conserved nat-miRNAs Target MADS Box Protein mRNAs. Although the majority of the identified miRNAs are rice-specific, nine miRNAs representing three groups are conserved in at least nine monocots but not in *Arabidopsis* and *Medicago* (Figs. 2 and 3 *C* and *D*). All of these miRNAs were predicted to target genes encoding MADS box proteins. We verified their function by identifying cleavage sites specifically mapping to the miRNA complementary sites (Fig. 1*C*). Several attributes of this group of miRNAs make them unusual compared with most known miRNAs. First, the mature miRNAs were derived from an unannotated but overlapping transcript antisense to the target MADS box genes (SI Fig. 7); hence, we called these nat-miRNAs for "natural antisense transcript miRNAs." These antisense transcripts were represented in EST collections. Second, the pre-miRNAs are processed from long transcripts with large introns. The hairpin structure could be predicted only after removal of the introns but not from the genomic sequence (Fig. 3*A* and SI Fig. 8). Third, more than one

PLANT BIOLOGY

Table 1. Identified rice miRNAs

miRNA family	Sequence (5′→3′)	Loci	Cons.†	Abund.‡	miR*	RNA blot	Predicted targets§	Proteins
miR820¶	UCGGCCUCGUGGAUGGACCAG	3	N	4,672	Y	Y	**Os03g02010**	Cytosine methyltransferase
miR1423	AGCGCCCAAGCGGUAGUUGUC	1	N	3,606	Y	Y	Os01g59560	Protein kinase
miR1424	AUGCACACUGAUGCUGAUUGU	1	N	1,108	N	Y	Os10g31030	MT-A70 family protein
							Os06g07770	
miR1425	UAGGAUUCAAUCCUUGCUGCU	1	N	1,002	Y	Y	Os08g01650	Rf1 proteins
							Os08g17970	
							Os10g35436	
							Os10g35640	
							Os10g35240	
miR1426	AGAAUCUUGAUGAUGAUUAAA	1	N	774	Y	Y	Os11g36530	DnaJ domain containing protein
miR1427	UGCGGAACCGUGCGGUGGCGC	1	N	731	Y	Y	Os04g32560	clpA homolog CD4B
miR1428	CGUUUUGCAAAUUCGCAGGCC	1	N	615	N	Y	Os03g43390	Transcription factor (LRR)
miR1429	GUUGCACGGGUUUGUAUGUUG	1	N	477	Y	Y	Os09g24430	Hypothetical protein
miR1430	UGGUGAGCCUUCCUGGCUAAG	1	N	336	Y	Y	Os10g20090	Cellulose synthase protein
miR1431	UUUGCGAGUUGGCCCGCUUGC	1	N	269	Y	Y	Os10g02970	Transcription factor (LRR)
miR1432	AUCAGGAGAGAUGACACCGAC	1	N	263	Y	Y	Os03g59770	EF hand family proteins
							Os03g59790	
							Os03g59870	
miR810-3p	UGAACACCGAUAUGCGUCAUC	1	N	171	Y	Y	Os02g03930	PPR protein
							Os01g65990	
miR1433	UGGCAAGUCUCCUCGGCUACC	1	N	157	Y	Y	Os03g62150	DnaJ domain containing protein
							Os04g40200	HNH endonuclease protein
miR444a-f.2	UGCAGUUGUUGUCUCAAGCUU	2	Y	50,434	Y	Y	**Os02g36924**	ANR1-like MADS box protein
	UGCAGUUGCUGCCUCAAGCUU	3		3,397			**Os02g49840**	
	UGCAGUUGUUGCCUCAAGCUU	1		279				
							Os08g33488	
miR444b,c.1‖	UGUUGUCUCAAGCUUGCUGCC	2	Y	1,649	Y	Y	Os02g36924	ANR1-like MADS box protein
							Os04g38780	
miR444d.3	UUGUGGCUUUCUUGCAAGUUG	1	Y	683	Y	Y	**Os02g49840**	ANR1-like MADS box protein

†Conserved between rice and maize.
‡Sum of abundance from all six libraries.
§Predicted targets with a score of 2.5 or less are listed. Targets validated by 5′ RACE are in bold. Different families have different target sites.
¶osa-miR820 and -miR810 were identified by a recent study (18). However, the 5′ end of the mature miRNA820 is different from the annotation in the Sanger database. miR810-3p is generated from a different end of the stem from the annotated miR810.
‖In the Sanger database, one miR444 sequence was mapped to the rice genome (miR444a) (Fig. 3*A*), although we did not identify an MPSS tag for it. Sunkar *et al*. also predicted and validated miR444-directed cleavage of the MADS box gene Os02g49840 (19). We found five additional genomic loci can form similar stem–loop structure like the miR444a precursor. In this study, all the nat-miRNAs are generated from these loci.

functional nat-miRNA species was generated from each individual locus, probably because of the long dsRNA stem. For example, as shown in Fig. 3*A*, nat-miR444d.2 and nat-miR444d.3 were generated in phase from the 3′ arm of the fold-back structure. Both are validated to target the mRNAs encoding MADS box proteins. Similarly, nat-miR444b.1 and -2 and nat-miR444c.1 and -2 reside in the same stem-loop and have an overlapping region of 15 nt. The miRNA* sequences were detected for three of them (Fig. 3*A*), indicating that these smRNAs originate from the stem-loop by Dicer activity. Although for most miRNAs, variants of the most abundant read were detected, nat-miR444.1 and nat-miR444.2 are unlikely to be the products simply of DCL1 slippage; both nat-miR444.1 and nat-miR444.2 were expressed at high levels, with small RNA sequences ±4 nt on each side found at much lower abundances. Therefore, the nat-miR444 hairpin structure may have two preferred and mutually exclusive sites for Dicer cleavage. Consistent with their origins in a common precursor, nat-miR444b.2 and nat-miR444b.1 and nat-miR444d.2 and nat-miR444d.3 have similar expression patterns (Fig. 3*B*). It is relevant to note that one miR444 species miR444a.1, which requires splicing and is conserved in monocots, was cloned in an early study (19) (Fig. 3*A* and SI Fig. 9).

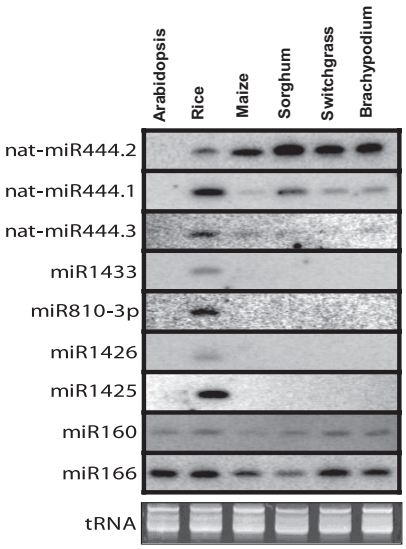

Fig. 2. Gel blots to evaluate the conservation of identified miRNAs in different plants. LMW RNAs isolated from seedlings of various plants were probed with labeled oligonucleotides.

Although nat-miRNA-generating loci resemble the nat-siRNA loci, several pieces of evidence strongly indicate that these small RNAs are indeed miRNAs and not siRNAs. For nat-siRNAs, the overlapping SRO5 and P5CDH antisense transcripts create a long (≈760-bp) double-stranded region, apparently processed by DCL2 to produce a unique 24-nt nat-siRNA (9). First, unlike nat-siRNA loci at which siRNAs are evenly distributed on both strands, >99% of the total small RNAs in the nat-miRNA loci mapped to the strand producing the mature miRNA, supporting a single-stranded

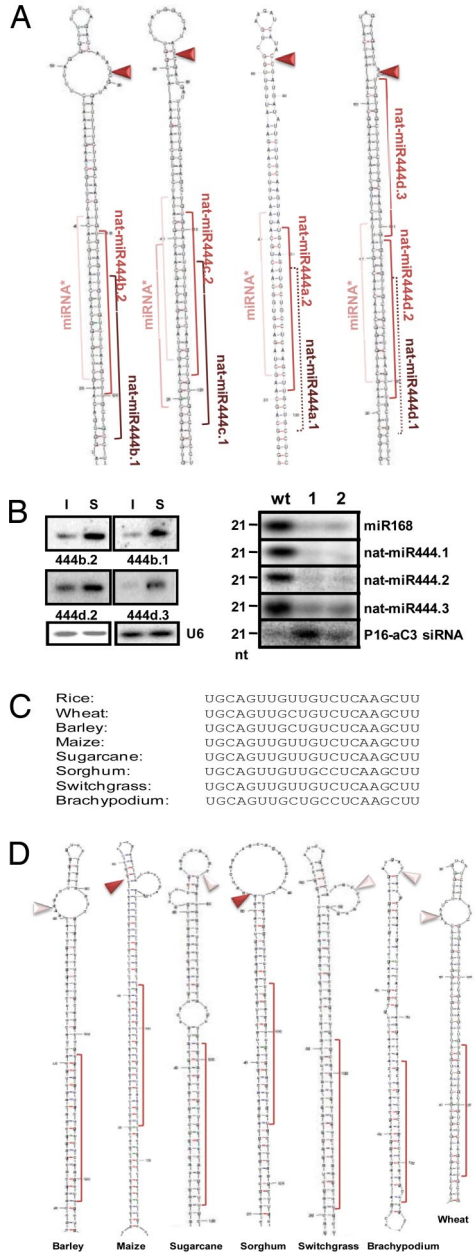

Fig. 3. nat-miRNA analysis. (*A*) Potential secondary structure of nat-miRNA precursors. The triangle indicates the ≈3-kb intron in the primary transcript. (*B*) Expression of nat-miRNAs. I, inflorescence tissues; S, seedlings; wt, wild-type rice seedlings; 1 and 2, seedlings from two independent DCL1 RNAi lines. The P16-aC3 siRNA is the same as described in ref. 33. (*C*) Alignment of nat-sm1 sequence from rice with the predicted homologs in other monocots. (*D*) Potential nat-sm1 precursors from other monocots. Orange triangles indicate the confirmed introns in maize and sorghum. Gray triangles indicate putative splicing sites in other precursors.

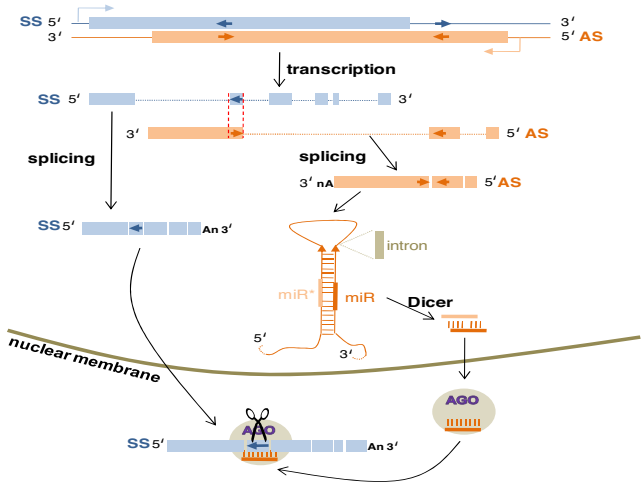

Fig. 4. Model for nat-miRNA biosynthesis and function. The nat-miRNA pathway initiates with the splicing of pri-miRNA transcripts to yield pre-miRNA hairpins. Our data support the contention that the possession of the introns in nat-miRNA precursors is important for the biosynthesis of miRNAs. Splicing of these introns limits the potential base-pairing of the pre-nat-miRNA with the sense transcript and favors hairpin formation. After Dicer cleavage, the mature nat-miRNAs then enter the cytoplasm and direct the cleavage of the sense transcripts that are their targets. AS, antisense strand; SS, sense strand.

origin. Second, although the spliced miRNA precursor can form an extensive stem-loop structure, >95% of small RNAs corresponded precisely to the miRNA and miRNA*. In contrast, siRNAs usually are distributed in a random or phased fashion along the dsRNA regions. Third, like canonical miRNAs, nat-miRNAs require DCL1 activity. As shown in Fig. 3*B*, all of the nat-miRNAs that we tested were greatly reduced in rice DCL1 RNAi transformants. In contrast, the production of tested endogenous siRNAs was not affected (Fig. 3*B*), which is clearly different from nat-siRNAs that depend on DCL2 but not DCL1. Fourth, the nat-siRNA sense–antisense gene pair has an extensive overlapping region. However, at the RNA level, the spliced nat-miRNA precursors and the mature MADS transcripts only share a very small overlapping region (≈65 nt) so that the thermodynamic rules do not favor the intermolecule interactions (dsRNA formation) over the intramolecule interactions (stem-loop formation) (Fig. 4 and SI Fig. 8). The nonextensive nat-miRNA and target pairing and the presence of introns minimize the chance of siRNA production. Although the genomic sequences for most other monocots are very limited, we found potential nat-miRNA precursors in sorghum and maize containing intronic sequences (Fig. 3*D*), in contrast to the previous report on miR444 (19). Putative splicing sites can be found in other precursors as well. Therefore, this kind of structure is probably important for nat-miRNA function and evolution. Fifth, because of the perfect complementarity between nat-miRNA and some of their targets, the nat-miRNA should target MADS transcripts for endonucleolytic cleavage (Fig. 4), which is indeed indicated by our 5′ RACE assays. The evidence supporting nat-miRNA-initiated cleavage is strong, because targets were cleaved at the canonical position for 93% of the sequenced clones (Fig. 1*C*). Considering this preponderance of evidence, these small RNAs are clearly nat-miRNAs, not nat-siRNAs. Our model for their production is illustrated in Fig. 4.

The identification of nat-miRNAs raises the intriguing question of their evolution. Because all of the nat-miRNA precursors can form long near-perfect fold-back structures, it is unlikely that these evolved by random mutation, and an inverted duplication origin seems more likely (24). Such a duplication could occur by transposon-mediated events (SI Fig. 10). If expressed, the inverted duplication segments could generate siRNAs by DCL processing of the fold-back structure from a single transcript or long dsRNA molecules from a sense–antisense transcript pair. The young nat-miRNA locus could evolve by steps including sequence divergence, adaption of the hairpin transcript to the miRNA biogenesis pathway, and additional duplication of the gene family (Fig. 4 and SI Fig. 10). The nat-miRNA loci are unusual in that an additional step, acquisition of intron splicing, appears essential to the formation of functional miRNA loci. For these nat-miRNAs, splicing clearly evolved before the amplification of the gene families, because it is shared among all of the nat-miRNA loci, some of which are not from the antisense strand of their targets (nat-miR444e and nat-

miR444f). It should be noted that several annotated *Arabidopsis* miRNAs (miR840, miR860, and miR408) overlap the 3′ UTR of other protein coding genes (11, 15, 28). However, the functional relationship between these miRNAs and the sense transcripts is not clear. None of them has been shown to target any mRNA for cleavage including the antisense genes. Therefore, it is likely that these miRNA genes evolved independently of the overlapping genes at these loci.

In many eukaryotes, up to 20% of genes have *cis*-antisense overlapping pairs (29) so there is considerable potential for some of these to generate nat-siRNAs or nat-miRNAs (9, 10). The nat-miRNAs, target sites and sense–antisense transcript arrangement that we discovered are conserved among monocots, indicating this pathway is at least 50 million years old. Yet, genome-wide analysis indicated that small RNAs are not preferentially associated with overlapping transcripts (14, 30), so identifying the loci that give rise to nat-small RNAs will be a future challenge. In the case of nat-miRNAs, the different intron/exon structures that we observed in the sense versus antisense transcripts minimized the overlap between the RNAs and were a prerequisite for hairpin precursor formation (Fig. 4). These findings suggest that filters based on these characteristics may identify additional nat-miRNAs that greatly influence the evolution and physiology of plants and other organisms.

Materials and Methods

Plant Growth. Nipponbare plants were grown in Conviron growth chambers with 12-h light/dark cycles, 80% relative humidity and 28°C during the day, and 60% and 25°C during the night. Inflorescence library was made from immature panicles from 90-day-old plants. Stem tissue was collected from 10-week-old plants. For ABA treatment, 21-day-old seedlings were sprayed with 100 μM ABA or water (control). Leaf tissue was collected after 6 h. For the biotic stress libraries, 21-day-old Nipponbare plants were challenged with rice blast (*M. grisea*) isolate Che086061. We counted the *M. grisea* spores under a microscope on a hemocytometer and prepared a spore suspension of 1×10^5 spores per ml. Then, we added three drops (200 μl) of Triton X-100 detergent to 50 ml of spore suspension to make sprayed spore suspension stay firmly on the rice leaf surface. Leaf tissue was harvested from water-sprayed control and pathogen-infected rice plants 48 h after inoculation. Tissues were snap-frozen in liquid nitrogen and stored at −80°C until small RNA was isolated. More information about the MPSS libraries can be found at http://mpss.udel.edu/rice.

MPSS Data Generation and Analysis. The small RNA libraries were constructed as described in ref. 14. The signatures were matched against The Institute for Genomic Research (TIGR) rice genome (version 4.0) and assigned to each location at which a perfect match was found. For the small RNA abundances, we merged the sequencing runs and calculated a single abundance normalized to TPQ after the removal of rRNAs and tRNAs signatures.

RNA Analysis. RNA blot hybridization analysis was performed as described in ref. 14, using total RNA extracted with TRIzol (Invitrogen). Low-molecular-weight (LMW) RNA gel blots, radiolabeled probes for specific small RNAs, and hybridization/wash conditions were described previously (14). All blots shown are representative of at least two independent experiments.

To detect small RNAs in wild-type and OsDCL1 RNAi plants (17), splinted ligation was performed by using miRtect-IT miRNA labeling and detection kit (USB) as described in ref. 31. For each small RNA, we designed specific bridge oligonucleotides as shown in SI Fig. 8 (32). By using 1 μg of total RNA, small RNAs were captured by specific bridge oligonucleotide and ligated to P^{32}-labeled detection oligo with T4 DNA ligase. Ligated products were separated on 15% urea-polyacrylamide gel and visualized with a phosphorimaging system.

RNA Ligase-Mediated (RLM) 5′ RACE. To map the cleavage site of target transcripts, a modified procedure for RLM 5′ RACE was performed by using the FirstChoice RLM-RACE Kit (Ambion). Total RNAs (1 μg) from 2-week-old seedling and inflorescence tissues were ligated to 5′ RACE adapter without calf intestine alkaline phosphatase treatment. The gene-specific outer primers then were used for cDNA synthesis. Initial PCR was carried out by using the 5′ RACE outer primer and gene-specific outer primer. Nested PCR was carried out by using 1/50 of the initial PCR, the 5′ RACE inner primer, and gene-specific inner primer. The gene-specific primers were listed in SI Table 5. RACE fragments were cloned into pGEM T-easy vector (Invitrogen) and sequenced after gel purification.

ACKNOWLEDGMENTS. We thank Mayumi Nakano for her work on the web interface; Dr. R. C. Venu and Dr. Guo-liang Wang (Ohio State University, Columbus, OH) for some of the rice samples; Dr. Susan Barton (University of Delaware, Newark, DE) for the switchgrass sample; Dr. David Garvin (US Department of Agriculture, Saint Paul, MN) for the Brachypodium sample; and Dr. Hajime Sakai and Randy Parker for their assistance in growing and caring for the rice plants. This work was supported by National Science Foundation Grant 0439186 (to P.J.G. and B.C.M.), U.S. Department of Agriculture Grants 2007-01991 (to P.J.G.) and 2005-35604-15326 (to B.C.M. and P.J.G.), with additional support from National Science Foundation Award 0321437 (to B.C.M.) and U.S. Department of Agriculture Agricultural Research Service Current Research Information System Project 1907-21000-014 (to D. Weir). D.-H.J. was partially supported by a Korean Research Foundation Fellowship funded by the Korean Government (KRF-2006-214-C00085), and S.R.T. was supported by National Institutes of Health Training Grant T32 GM-08550.

1. Chen X (2005) MicroRNA biogenesis and function in plants. *FEBS Lett* 579:5923–5931.
2. Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol* 57:19–53.
3. Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes Dev* 20:759–771.
4. Meins F, Jr, Si-Ammour A, Blevins T (2005) RNA silencing systems and their relevance to plant development. *Annu Rev Cell Dev Biol* 21:297–318.
5. O'Donnell KA, Boeke JD (2007) Mighty Piwis defend the germline against genome intruders. *Cell* 129:37–44.
6. Vazquez F, *et al.* (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell* 16:69–79.
7. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. *Plant J* 46:243–259.
8. Zhang X, *et al.* (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126:1189–1201.
9. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 123:1279–1291.
10. Katiyar-Agarwal S, *et al.* (2006) A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci USA* 103:18002–18007.
11. Fahlgren N, *et al.* (2007) High-Throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of miRNA genes. *PLoS ONE* 2:e219.
12. Henderson IR, *et al.* (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* 38:721–725.
13. Lu C, *et al.* (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res* 16:1276–1288.
14. Lu C, *et al.* (2005) Elucidation of the small RNA component of the transcriptome. *Science* 309:1567–1569.
15. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev* 20:3407–3425.
16. Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 14:787–799.
17. Liu B, *et al.* (2005) Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. *Plant Physiol* 139:296–305.
18. Luo YC, *et al.* (2006) Rice embryogenic calli express a unique set of microRNAs, suggesting regulatory roles of microRNAs in plant post-embryogenic development. *FEBS Lett* 580:5111–5116.
19. Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and characterization of microRNAs from rice. *Plant Cell* 17:1397–1411.
20. Nobuta K, *et al.* (2007) An expression atlas of rice mRNAs and small RNAs. *Nat Biotechnol* 25:473–477.
21. Vaucheret H, Mallory AC, Bartel DP (2006) AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. *Mol Cell* 22:129–136.
22. Xie Z, *et al.* (2005) Expression of Arabidopsis MIRNA genes. *Plant Physiol* 138:2145–2154.
23. Wang XJ, Reyes JL, Chua NH, Gaasterland T (2004) Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biol* 5:R65.
24. Allen E, *et al.* (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat Genet* 36:1282–1290.
25. Akagi H, *et al.* (2004) Positional cloning of the rice Rf-1 gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. *Theor Appl Genet* 108:1449–1457.
26. Wang Z, *et al.* (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* 18:676–687.
27. Boonburapong B, Buaboocha T (2007) Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins. *BMC Plant Biol* 7:4.
28. Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell* 16:2001–2019.
29. Wang XJ, Gaasterland T, Chua NH (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. *Genome Biol* 6:R30.
30. Henz SR, *et al.* (2007) Distinct expression patterns of natural antisense transcripts in Arabidopsis. *Plant Physiol* 144:1247–1255.
31. Maroney PA, Chamnongpol S, Souret F, Nilsen TW (2007) A rapid, quantitative assay for direct detection of microRNAs and other small RNAs using splinted ligation. *RNA* 13:930–936.
32. Sunkar R, Girke T, Zhu JK (2005) Identification and characterization of endogenous small interfering RNAs from rice. *Nucleic Acids Res* 33:4443–4454.